SPECIFICATION

Electronic Version 1.2.8

Stylesheet Version 1.0

[*NUCLEIC ACID SEQUENCING METHOD*]

Cross Reference to Related Applications

This application claims the priority benefit of Taiwan application serial no. 91118507, filed August 16, 2002.

Background of Invention

[0001] Field of Invention

[0002] The present invention relates to a nucleic acid sequencing method. More particularly, the present invention relates to a nucleic acid sequencing method for rapid sequencing by a rotating electric field. In the present invention, the nucleic acid includes deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

[0003] Description of Related Art

[0004] After the human genome project, it has become promising to use gene sequencing to diagnose or treat genetic diseases. Therefore, much research has been initiated to develop the methods and/or instrumentation for gene or nucleic acid sequencing.

[0005] The prior art nucleic acid sequencing method proposed by Fredrick Sanger is performed by replicating DNA under controlled conditions to obtain fragments of various lengths, so that the complete DNA sequence can be derived. The following paragraph details the prior art nucleic acid sequencing method.

[0006] Several polymerase chain reaction (PCR) reagents, including polymerase I, specific primers with complementary sequences, deoxyribonucleotide triphosphate (dNTPs) and buffers, are provided. The dNTPs are labeled with either isotopes or fluorescent

molecules. On the other hand, dNTP analogs are prepared. The dNTP analogs lack the 3"-hydroxyl groups for forming the phosphodiester bond with the next subunit, thus terminating the elongation of DNA. After four different dNTP analogs are added into four groups of PCR reagents respectively, four groups of chain terminated fragments finished with four dNTP analogs are obtained through PCR reactions. Later on, electrophoresis is used to separate DNA fragments with diversified and different lengths. The DNA fragments are detected either through isotopes or fluorescent molecules. By comparing the sizes (or the positions and spacing in the gel) of each fragment, the DNA sequence is obtained.

[0007] Although DNA sequencing has important medical applications, so far the prior art methods in sequencing nucleic acid (or polynucleotides) are time-consuming, costly, and inaccurate. For example, the prior Sanger method used for human genome sequencing took 15 years and cost nearly 3 billion USD. Not only are the PCR reaction and electrophoresis analysis very time-consuming, but the instrumentation and reagents are also very expensive. Because of its slowness and high-price, the prior art DNA sequencing technique is impractical in diagnosing diseases, especially acute or epidemic diseases.

[0008] Moreover, accuracy problems exist in the prior art DNA sequencing method, as shown in *Cell, 106, 413 (2001)*. A comparison of Celera and Ensembl predicted gene sets reveals 20% overlap in novel genes. It is expected that the similarity of uncoded regions between these two groups" results is even smaller. Therefore, it is highly desirable to develop an accurate nucleic acid sequencing method that is fast and inexpensive.

Summary of Invention

[0009] The invention provides a nucleic acid sequencing method, which is time-efficient and low-cost.

[0010] The invention provides a fast single-molecule nucleic acid sequencing method with higher accuracy.

[0011] As embodied and broadly described herein, the present invention provides a nucleic acid sequencing method. After providing a thin film with a nanopore and

placing the thin film in a buffer solution, nucleic acid sequences are added into the buffer solution. The nucleic acid sequence can be a DNA sequence or an RNA sequence. An applied electric field perpendicular to the thin film drives the nucleic acid sequence to pass through the nanopore of the thin film. At the same time, a rotating electric field parallel to the thin film is applied to control the movement (i.e. the translocation speed) of the nucleic acid sequence through the nanopore. The rotating electric field controls whether the nucleic acid sequence is stretched or unstretched. The frequency of the rotating electric field correlates to the translocation speed of the nucleic acid sequence through the nanopore. For a rotating electric field with high frequency, the polynucleotide sequence will rapidly pass through the nanopore. On the other hand, for a rotating electric field with low(er) frequency, the translocation speed of the polynucleotide sequence is under control. With an adequate frequency, the rotating electric field can control only one nucleotide of the polynucleotide sequence passing through the nanopore at a time. Since different nucleotides (i.e. A, G, T, C) cause different levels of blockage toward the nanopore, measured blockage currents for different kind of nucleotides are distinct. In the present invention, an outer circuit is applied to measure the blockage currents and change of blockage currents over time, so that the polynucleotide sequence can be determined by measuring the change of blockage currents over time. The method of the present invention further includes adding two extra fragments at both ends of the tested sequence respectively. These two fragments can be used to label different ends (3" or 5" end) and to locate the main sequence.

[0012]

The sequencing method of the present invention can be performed in the array format by forming array cells in the thin film and forming one nanopore in each cell. Such sequencing array design is useful in making comparison for different sets of results from the same nucleic acid sequence. Sincethe translocation time of each nucleotide is in multiples of $T_c/4$ (time for the sequence to remain stretching) and each kind of nucleotide has a distinct blockage current, numbers of repeating nucleotides in sections of the polynucleotide sequence can be obtained by comparing the measured time of a specific section (in unit of $T_c/4$) to obtain the smallest integer multiple. The change of blockage currents over time of the same nucleic acid sequence is measured several times, in order to obtain different sets of results. By

making comparisons between different sets of results from the same sequence, prediction errors can be greatly reduced and the polynucleotide sequence can be determined accurately.

[0013] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and are intended to provide further explanation of the invention as claimed.

Brief Description of Drawings

[0014] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention. In the drawings,

[0015] Fig. 1 is a display view of a nucleic acid sequence passing a nanopore of a thin film according to one preferred embodiment of the present invention;

[0016] Fig. 2 is a display view of a nucleic acid sequence passing a nanopore of a thin film under the influence of a rotating electric field and a stable electric field according to one preferred embodiment of the present invention;

[0017] Fig. 3 is a display view showing a simulated polynucleotide sequence passing through the nanopore according to the bond-fluctuation model in a cubic lattice;

[0018] Fig. 4 is a display view showing a simulated polynucleotide sequence passing through the nanopore according to the off-lattice bead-spring model;

[0019] Figs. 5A and 5B show the relationship of the numbers of nucleotides that have passed the pore versus time in the translocation processes under high frequencies and low frequencies, respectively;

[0020] Fig. 6 shows quantization of the translocation time for each nucleotide passing through the nanopore under low frequency rotating electric field, while the nucleotides are marked according to the lengths of their translocation time, not their order in the sequence;

[0021] Fig. 7 shows the relationship of the translocation time for the polynucleotide

sequence passing through the pore versus the frequency of the rotating electric field. The inset shows the translocation time of each nucleotide of the polynucleotide during a typical translocation process simulated by the off-lattice bead-spring model;

[0022] Fig. 8A shows the relationship of the number of nucleotides passing through the nanopore versus time, while Fig. 8B illustrates the relationship between the translocation time of the nucleotides in Fig. 8A and their measured blockage currents; and

[0023] Fig. 9 shows the relationship of the prediction errors versus the number of the measurements.

Detailed Description

[0024] Fig. 1 is a display view of a nucleic acid sequence passing a pore of a thin film according to one preferred embodiment of the present invention.

[0025] As shown in Fig. 1, a membrane or thin film 100 is provided with a nanopore 102. The thin film 100 is made of, for example, silicon nitride. For example, an ion beam is used to form the nanopore 102 and the nanopore 102 has a size of about 2-3 nm. For the method for forming nanopores in the thin film refer to J. Li et al., *Nature* 412, 166 (2001).

[0026] After the thin film 100 is placed in a buffer solution 106 (as shown in Fig. 1), nucleic acid sequences 104 are added into the buffer solution 106. The nucleic acid sequence 104 can be a DNA sequence or an RNA sequence, and sometimes is denoted as a polynucleotide sequence or a polynucleotide. Since the nucleic acid sequence 104 is a long chain with negative charges, an applied electric field perpendicular to the thin film can drive the nucleic acid sequence 104 to pass through the nanopore 102 of the thin film 100.

[0027] As shown in Fig. 2, the nucleic acid sequence 104 is driven by a uniform-amplitude electric field E in the z-direction to pass through a nanopore of size D in the thin film 100. A rotating electric field E_c on the x-y plane is added in order to control the movement (i.e. the translocation speed) of the nucleic acid sequence 104 through the nanopore 102.

[0028] Therefore, except for the applied electric field E perpendicular to the thin film that drives the nucleic acid sequence 104 to pass through the nanopore 102 of the thin film 100, the rotating electric field E_c parallel to the thin film 100 controls stretching or unstretching (releasing) the long-chain nucleic acid sequence 104 so as to control the translocation speed of the nucleic acid sequence 104 through the nanopore 102.

[0029] According to the preferred embodiment, the rotating electric field E_c is formed by one set of parallel electrode pairs perpendicular to another set of parallel electrode pairs. One set of parallel electrode pairs generates a sinusoid (sine) AC electric field, while the other set of parallel electrode pairs generates a cosinusoid (cosine) AC electric field. With the same frequency, the combination of these two electric field having a 90-degree phase difference forms a circular rotating electric field $E_c = E_c \sin(\omega t)\, i + E_c \cos(\omega t)\, j$, while i and j are unit vectors in the x- and y-directions, as shown in Fig. 2.

[0030] The rotating electric field E_c controls whether the nucleic acid sequence 104 is stretched or unstretched. If the nucleic acid sequence 104 is fully stretched by the rotating electric field E_c, the nucleic acid sequence 104 above the nanopore 102 cannot travel through the nanopore 102. Only if the nucleic acid sequence 104 becomes relaxed (unstretched) by the rotating electric field E_c, can the nucleic acid sequence 104 above the nanopore 102 travel through the nanopore 102. The frequency ω of the rotating electric field E_c correlates to the translocation speed of the nucleic acid sequence 104 through the nanopore 102. For a rotating electric field with high frequency, the polynucleotide sequence rapidly passes through the nanopore. On the other hand, for a rotating electric field with low(er) frequency, the translocation speed of the polynucleotide sequence is controlled. With an adequate frequency, the rotating electric field can be controlled to only one nucleotide of the polynucleotide sequence passing through the nanopore 102 at a time. Moreover, the translocation time (i.e. the time required for passing through the nanopore) of each nucleotide is found to be $mT_c/4$, where m is an integer and $T_c/4$ is the time that the sequence remains stretching, while T_c is the period of the rotating electric field.

[0031] Figs. 3 and 4 are display views showing a simulated polynucleotide sequence passing through the nanopore.

[0032] Referring to Figs. 3 and 4, a polynucleotide sequence 104 of length N (i.e. having N nucleotides) is represented by the bond-fluctuation model in a cubic lattice (as shown in Fig. 3) or the off-lattice bead-spring model (as shown in Fig. 4). In the preferred embodiment, the polynucleotide sequence 104 is a single strand DNA (ssDNA), and the Metropolis Monte-Carlo (MC) algorithm at a constant temperature T is used to simulate its motion.

[0033] In the bond-fluctuation model, each nucleotide occupies a cube of length 1 (lattice spacing) and the set of allowed bond vectors is $B = P(2,0,0) \cup P(2,1,0) \cup P(2,1,1) \cup P(2,2,1) \cup P(3,0,0) \cup P(3,1,0)$, where $P(a, b, c)$ stands for the set of all permutations and sign combinations of $\pm a, \pm b, \pm c$. This model has been shown to be a realistic and efficient method for studying polymer dynamics in various systems and has been cited in a few references, including C.-M. Chen, Y.-A. Fwu, *Phys. Rev.* E63, 011506(2001).

[0034] Referring to Figs. 2-4, in the simulations, the nucleic acid sequence (polynucleotide sequence) 104 is driven by the uniform electric field E in the z-direction (perpendicular to the thin film 100) to pass through the nanopore 102 of the thin film 100. Above the thin film, the electric field E_c rotates on the x-y plane (parallel to the thin film 100) to control the translocation speed of the polynucleotide 104 passing the nanopore 102. Both the frequency and the amplitude of the rotating field can be used to control the movement (i.e. translocation speed) for each nucleotide of the polynucleotide sequence.

[0035] At each instant, a nucleotide is picked up at random and attempts to move in any of the six directions by one lattice spacing. If any attempted move of nucleotides satisfies the excluded volume constraint and the new bond vectors are still in the allowed set, the move is accepted with probability $p = \min[1, \exp(-\Delta U / kT)]$, where ΔU is the energy change of the chain and kT is thermal energy. In this model, the energy of polynucleotide sequence 104 is expressed as $U = U_{\text{bend}} + U_{\text{electric}} + U_{\text{H-bond}}$, where $U_{\text{bend}} = \Sigma_i e(1 \cos \theta_i)$ is the bending energy of the chain with a rigidity e and a bending angle θ_i. U_{electric} is the electric potential energy due to a constant electric field in the z- direction and a rotating electric field on the $x-y$ plane, and $U_{\text{H-bond}}$ is the hydrogen bonding energy of (A, T) and (G, C) pairs. Here it

is considered to have negligible hydrogen bonding between bases, which can be realized by adjusting pH value, raising temperature, or adding urea.

[0036] To study the kinetics of the polynucleotide 104 passing through the nanopore 102, each set of parameters for the translocation process is simulated 50 times. For a chain of 50 nucleotides, we choose the pore size $D = 3$, temperature $T = 1$, the uniform electric field amplitude $E = 1.5$, and the bending rigidity $e = 0.2$. Here thermal energy and electric charge of each nucleotide are set to unity, and the corresponding electric field is of order 10^{7} V/m. For the rotating field, its frequency ω varies from 10^{-1} to 10^{-8} (MC step^{-1}) and its amplitude E_c varies from 0.1 to 1.2.

[0037] Figs. 5A and 5B show the relationships of the numbers of nucleotides that have passed the pore versus time in the translocation process under high frequencies and low frequencies, respectively. As shown in Fig. 5A, at high frequencies ($\omega \geq 10^{-3}$), the polynucleotide sequence passes through the pore smoothly and the translocation time of the whole chain, t_c, is about a constant ($t_c \sim 2 \times 10^{4}$ MC steps). In these cases, the translocation time of each nucleotide, t_n, does not vary dramatically.

[0038] As shown in Fig. 5B, at low frequencies ($\omega \leq 10^{-4}$), two kinds of translocation kinetics are observed for the polynucleotide sequence. For nucleotides located at the middle of the sequence, t_n is much longer than that of nucleotides near both ends of the sequence. At $\omega = 10^{-6}$, t_c is about 10^{8} MC steps. It has been estimated previously that t_n is about 1 microsecond at the present driving electric field strength for a smooth translocation and thus 1 MC step in the simulation is in the order of 10^{-8} sec. It is concluded that the frequency of the rotating field should be less than 10^{4} Hz in order to slow down the translocation process.

[0039] Fig. 6 shows quantization of the translocation time for each nucleotide passing through the nanopore under a low frequency rotating electric field, while the nucleotides are marked according to the lengths of their translocation time, but not their order in the sequence.

[0040] As shown in Fig. 6, the two axes are nucleotide versus the translocation time t_n each nucleotide. The detailed study reveals that $t_n \approx mT_c/4$ for $\omega \leq 10^{-4}$ and E_c /

$E > 0.4$, where m is an integer and $T_c = 2\pi / \omega$. This effect of quantized t_n can be explained as follows. At some instant, the remaining segment of the chain (the polynucleotide sequence) above the thin film is aligned along the direction of the rotating electric field. In this case, the chain is taut and the nucleotide near the nanopore cannot pass through the pore. The stretched polynucleotide chain does not move with the rotating field due to lattice effects when the rotating field rotates away from the aligned direction. When the rotating field is almost perpendicular to the aligned direction, the stretched chain starts to move and becomes loose (unstretched). If the response of the chain is faster than the rotation of the rotating electric field, it will be quickly aligned along the new direction of the rotating field again. Since the nucleotide near the pore can pass through the pore only when the chain is loose, t_n must be a multiple of $T_c/4$. Deviation of t_n from predicted values would depend on the stability of the stretched chain and its response toward the rotating electric field.

[0041]

Fig. 7 shows the translocation time for the polynucleotide sequence passing through the pore versus the frequency of the rotating electric field. In order to eliminate possible lattice effects, off-lattice simulations of the polynucleotide translocation process are also carried out. The inset of Fig. 7 shows the translocation time of each nucleotide of the polynucleotide in a typical translocation process using the off-lattice bead-spring model, in which the quantization of t_n is clear. The simulated polynucleotide sequence in the off lattice bead-spring model behaves differently from that in the cubic lattice bond-fluctuation model. The polynucleotide sequence in the off lattice model cannot pass through the thin film as a whole under low frequency rotating electric field. However, by transitorily shutting off the rotating electric field or tuning the rotating electric field to a higher frequency (the shutting time is 0.02 period for every 1/4 period of the rotating electric field in the inset), the polynucleotide sequence in the off lattice bead-spring model can behave in the same way as the sequence in the cubic lattice bond-fluctuation model, as shown in the inset. Referring to Fig. 7, the two axes are the translocation time t_c of the polynucleotide sequence versus the frequency ω of the rotating electric field, showing the dependence of t_c on ω. If the response of the chain is too slow, it will always be loose and penetrate the pore smoothly. Fig. 7 shows that t_c is inversely proportional to ω for $\omega \leq 10^{-4}$ and is almost a constant for $\omega \geq 10^{-3}$. The boundary between

these two regimes depends on the response of the polynucleotide sequence and can be varied by changing the viscosity of the solution or the friction of the thin film surface.

[0042] The aforementioned circumstances apply to the polynucleotide sequence of 30 nucleotides, 70 nucleotides and 100 nucleotides.

[0043] The present invention further includes linking two specific sequence fragments to both ends (the 3" end and 5" end) of the polynucleotide sequence, in order to tell the differences between both ends. In the preferred embodiment, we select a polynucleotide consisting of 26 randomly generated nucleotides (GTACTTCGCGTGTAGTCATTTAATCC) located at the middle and two extra fragments AAAAAAAAAAAC and ACCCCCCCCCCC attached at the 3" and 5" ends, respectively. These two fragments are added because nucleotides near both ends pass through the nanopore quickly and cannot be sequenced, as indicated in Fig. 5B. In addition, they can be used to locate the main sequence.

[0044] In Fig. 8A, the relationship of the number of nucleotides passing through the nanopore versus time is shown, while Fig. 8B illustrates the relationship between the translocation time of the nucleotides in Fig. 8A and their measured blockage currents. Since different nucleotides (i.e. A, G, T, C) cause different levels of blockage toward the nanopore, the measured blockage currents for different kind of nucleotides are distinct. In the present invention, an outer circuit is applied to measure the blockage currents and change of blockage currents over time, so that the polynucleotide sequence (i.e. the linking order of the nucleotides) can be determined by measuring the change of blockage currents over time. Using the aforementioned 26-nucleotide polynucleotide sequence having two extra fragments AAAAAAAAAAAC and ACCCCCCCCCCC attached at the 3" and 5" ends as an example, if the blockage currents of A, G, T, C are assumed to be 18, 20, 35, and 40, an increase in the blockage current from 18 to 40 signals the beginning of the main sequence from the 3" end, while a drop of the blockage current signals the beginning of the sequence from the 5" end.

[0045] The sequencing method of the present invention can be performed in the array format by forming array cells in the thin film and forming one nanopore in each cell

using an ion beam, for example. The stable electric field and the rotating electric field are applied to control translocation of the polynucleotide sequence through the pore. Each of the sequencing array cells can be controlled and measured independently or in rows by applied circuits.

[0046] Such sequencing array design is useful in making comparison for different sets of results from the same nucleic acid sequence. Since t_n (translocation time of each nucleotide) is a multiple of $T_c/4$ (time for the sequence to remain stretching) and each kind of nucleotide has a distinct blockage current, the numbers of repeating nucleotides in sections of the polynucleotide sequence can be obtained by comparing the measured time of a specific section (in unit of $T_c/4$) to obtain the smallest integer multiple. The change of blockage currents over time of the same nucleic acid sequence is measured several times, in order to obtain different sets of results. By making comparisons between different sets of results from the same sequence, prediction errors can be greatly reduced and the polynucleotide sequence can be determined accurately.

[0047] As shown in Fig. 9, the two axes are the prediction errors and the number of the measurements (i.e. being measured for how many times). For one single measurement, the prediction error of the random sequence is about 30%. If more than 16 sets of results are used in analysis and comparison, the sequencing error is reduced to nearly zero. It is evident that the prediction error decreases rapidly as the number of measurements increases.

[0048] Note that, experimentally, accuracy of sequencing mainly relies on the differences of blockage currents for (A, G) or (C, T). Adding a chemical group to the specific base of the nucleotide can magnify the differences of blockage currents between different nucleotides. For example, a benzoyl group can be attached to the amino functional groups of A, G, and C.

[0049] In the simulations, a strong electric field is considered to be able to reduce thermal effects (for example, backward motion of nucleotides against the driving electric field), and these thermal effects become more significant if a weak electric field is applied. Nevertheless, prediction errors due to thermal effects can be reduced or even cancelled if many sets of measurement results are used for analysis.

[0050] In conclusion, the present invention has the following advantages:

[0051] 1.The present invention provides a rapid single-molecule nucleic acid sequencing method. If performed in the array format, the method of the present invention can determine up to 100 million bases per day.

[0052] 2.No other reagents or special enzymes are required for the method of the present invention, thus reducing the costs.

[0053] 3.The present invention provides an accurate nucleic acid sequencing method, in combination of the sequencing array cells. The sequencing error is reduced to nearly zero by analyzing and comparing numerous sets of results obtained from the array cells.

[0054] 4.According to the method of the present invention, a convenient, accurate and cheap sequencing system can be designed. Such a system can be used to diagnose diseases, in medical treatment or biological sample detection.

[0055] It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.